

PB88-202031

Geostatistical Analysis of
Palmerton Soil Survey Data

(U.S.) Environmental Monitoring Systems Lab.
Las Vegas, NV

1987



U.S. Department of Commerce
National Technical Information Service



TECHNICAL REPORT DATA (Please read Instructions on the reverse before completing)		
1. REPORT NO. EPA/600/J-87/349	2.	3. REPORT NO. PB88-202031
4. TITLE AND SUBTITLE GEOSTATISTICAL ANALYSIS OF PALMERTON SOIL SURVEY DATA (Journal Version)	5. REPORT DATE 1987	
	6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Thomas H. Starks, Allen R. Sparks, Kenneth W. Brown	8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Environmental Research Center, University of Nevada Las Vegas, Las Vegas, NV 89154, U.S. Environmental Protection Agency, EMSL-LV, Las Vegas, NV 89114	10. PROGRAM ELEMENT NO.	
	11. CONTRACT/GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring Systems Laboratory - LV, NV Office of Research and Development U.S. Environmental Protection Agency Las Vegas, NV 89193-3478	13. TYPE OF REPORT AND PERIOD COVERED	
	14. SPONSORING AGENCY CODE EPA/600/07	
15. SUPPLEMENTARY NOTES REF: Environmental Monitoring and Assessment, 9:239-261, 1987		
16. ABSTRACT This paper describes statistical and geostatistical analyses of data from a soil sampling survey. Soil sampling was performed, in October and November 1985, to obtain information on the level, extent, and spatial structure of metal pollution of the soil in and around the Palmerton, Pennsylvania, NPL Superfund site. Measurements of the concentrations of cadmium, copper, lead, and zinc in the soil samples were obtained. An appropriate variance stabilizing transformation was determined. Estimation of variance components was performed. Generalized covariance functions for log-transformed concentrations were estimated for each metal. Block kriging was employed using the estimated spatial structure models to obtain estimated metal concentration distributions over the central part of Palmerton. <i>→ PI</i>		
17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
REPRODUCED BY U.S. DEPARTMENT OF COMMERCE National Technical Information Service SPRINGFIELD, VA. 22161		
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 14
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE PC\$12.95/MF\$6.95

GEOSTATISTICAL ANALYSIS OF PALMERTON SOIL SURVEY DATA

THOMAS H. STARKS

Environmental Research Center, University of Nevada, Las Vegas, Las Vegas, NV 89154, U.S.A.

ALLEN R. SPARKS

P.O. Box 15027, Las Vegas, NV 89114, U.S.A.

and

KENNETH W. BROWN

*U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory,
P.O. Box 15027, Las Vegas, NV 89114, U.S.A.*

(Received 15 September, 1986)

Abstract. This paper describes statistical and geostatistical analyses of data from a soil sampling survey. Soil sampling was performed, in October and November of 1985, to obtain information on the level, extent, and spatial structure of metal pollution of the soil in and around the Palmerton, Pennsylvania, NPL Superfund site. Measurements of the concentrations of cadmium, copper, lead, and zinc in the soil samples were obtained. An appropriate variance stabilizing transformation was determined. Estimation of variance components was performed. Generalized covariance functions for log-transformed concentrations were estimated for each metal. Block kriging was employed using the estimated spatial structure models to obtain estimated metal concentration distributions over the central part of Palmerton.

1. Introduction

Soil samples were obtained to gather information about the level, extent, and spatial structure of concentrations of cadmium, lead, copper, and zinc in the soil in and around the Palmerton, Pennsylvania, NPL Superfund site (i.e., a site specified under CERCLA, the Comprehensive Environmental Response, Compensation and Liability Act of 1980). The sources of the metal contamination of the soil are alleged to be two zinc smelter operations, denoted herein as the East and West Plants, and a 4 km long slag pile located near Palmerton (see Figure 1). The West Plant started operations in 1898 and the East Plant began in 1914. The two plants underwent modification in 1980 that drastically reduced metal emissions.

The purpose of the soil survey was to obtain information for use in planning a more definitive survey of the metal concentrations in the soil near Palmerton to be used in planning remedial actions. The sampling was performed in October and November of 1985, by the firm of R.E. Wright Associates, Inc. This firm also dried, mixed, and sieved the samples and split out subsamples sent for laboratory analysis. These subsamples were analyzed using analytical method EPA-600-4-79-520, Metals

Environmental Monitoring and Assessment 9 (1987) 239-261.

© 1987 by D. Reidel Publishing Company.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

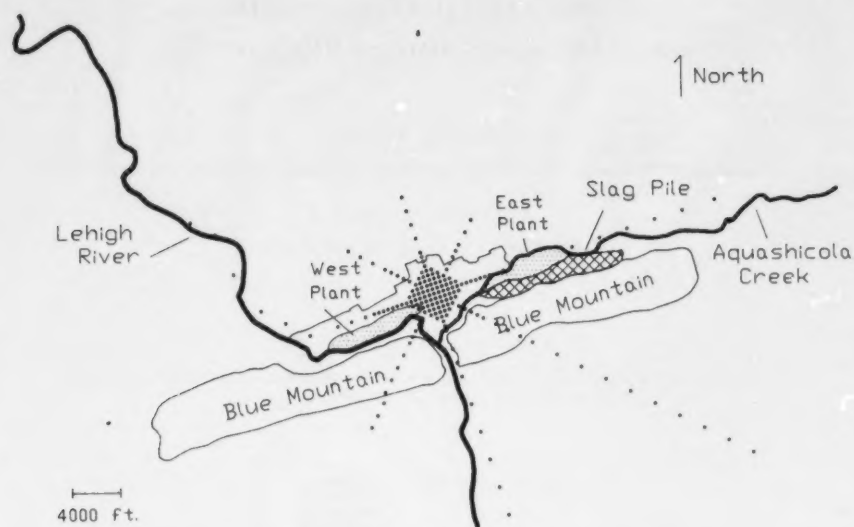


Fig. 1. Palmerton and designated sample point locations.

by Atomic Absorption Methods, by the Soil and Environmental Chemistry Laboratory of The Pennsylvania State University.

The town of Palmerton lies in a narrow valley bounded on the south by Blue Mountain and on the north by large hills and ridges. It is approximately 25 miles north of Allentown, Pennsylvania. The soils in and around Palmerton consists of unconsolidated deposits of glacial outwash. These deposits are poorly sorted, stratified sandy gravels with some interbedded red clay.

The purpose of this paper is to demonstrate how (and why) statistical and geostatistical techniques are combined in the analysis of soil survey data to locate, and estimate the size of important sources of variation, and to estimate spatial structure of pollutant concentrations in the soil. In particular, it is intended to show the need for variance stabilizing transformations, and for care in determining appropriate models for spatial structure so that the models will be consistent with the data and with the process that created the system under study.

2. Sample Design

The samples were collected at 88 points on a square grid over a diamond shaped area at the center of Palmerton and at 119 points along eight transects emanating out from the center of the grid to locations as far as 9.5 km from the center of the grid (see

Figure 1). (The transect to the west from Palmerton was bent to follow the Lehigh River valley). The designated distance between sample points on the square grid was 122 m (400 ft.). On the transects, the distances between sample points varied from 122 m up to 366 m (except where no samples were taken when a transect crossed smelter property). The original sampling plan called for 211 sample points, and only four points were not sampled. Failure to sample was either due to inability to get authorization from a land owner or to nature of terrain. Samples were not always taken at the designated sample point due to physical obstructions (i.e., rocks, highways, buildings, etc.). The field crew was allowed to move to the nearest unobstructed spot to take their sample provided that spot was within 61 m (200 ft.) of the designated point. The location where the actual sample was obtained was recorded and used in the spatial analysis of the data. The soil samples from 30 sample points on the transects were archived without measurement. Hence, measurements of metal concentrations were obtained at 177 points. A discussion of the rationale for the design is given in Starks *et al.* (1986).

At each sample point, a core was taken from each of the four major compass points on a 6 m diameter circle centered at the designated sample point. The depth of the core was 15 cm at all but 19 points where cores of 30 cm were obtained. The core diameter in each case was 1.9 cm (0.75 in). For all but ten sample points, the four cores from each sample point were composited. At ten sample points where 30 cm cores were taken, each core was cut into twelve 2.5 cm segments and the four segments from the same depth at a sample point were composited so that there were 12 samples to be analyzed from each of these ten sample points to provide information on how metal concentrations change with depth. For the other nine sample points where 30 cm cores were obtained, each core was cut into two 15 cm segments and segments from the same depth were composited to give two samples from each of these sample points for further investigation of depth effect. At ten sample points with 15 cm cores, the cores were not composited, but analyzed separately, so as to obtain information needed to estimate how much of the variation in metal concentrations between sample points was due to variation between core concentrations at individual sample points.

For quality assurance purposes, duplicate pairs (co-located samples) were obtained at ten sample points. This was done by first taking the usual sample of four 15 cm depth cores and compositing them, and then taking and compositing another sample of four 15 cm cores at locations within 0.5 m of the cores in the first sample. Analyses were performed on both samples and the differences in the results are due to short range spatial variation of the metal concentrations in the soil and also due to variation in sample taking, handling, subsampling, and analytical techniques.

Two subsamples (called splits) were taken, after drying, sieving, and mixing, from ten samples and sent separately to the laboratory for analyses. The results from splits gave information on the amount of variation in the results due to subsampling and analytical errors. Splits and duplicates were from different sample points.

3. Methods of Data Analysis

The first step in the data analysis consisted of plotting the measurements at their sample point locations on the map to visually inspect for trends, anomalies, and range of values. Next, quality assurance data from duplicate pairs were inspected to determine an appropriate transformation to stabilize the variance (i.e., to obtain data where the variance is not dependent on the mean concentration). After transformation, the variance of duplicates was calculated, to determine the contribution of short range spatial variation and variation in sample handling, subsampling, and analytical techniques to total variation between samples. Also, after transformation, variance between splits was obtained to determine the contribution of subsampling and analytical errors to total variance of the data. In addition, variance between cores within samples was estimated to find its contribution to total variance.

The spatial structures of the transformed concentrations of the four metals were obtained using a method consisting of cross-validation and response surface analysis (see Starks and Sparks, 1987). This procedure uses the fact that if the spatial structure model is correct, point kriging gives unbiased estimators of (transformed) concentrations and their estimation standard deviations. The cross-validation involves point kriging at sample points using the measurements at nearby neighbors to obtain standardized residuals (i.e., [observed value - predicted value]/estimated standard deviation). A criterion that measures the difference between the set of standardized residuals and the expected results in a sample from a standard normal distribution is used to compare different models and to provide direction in the search for a 'best' model.

Block kriging on $200' \times 200'$ blocks with the estimated models for spatial structure using the BLUEPACK software package, yields plots and isopleths of the average metal concentrations over the blocks within the region covered by the square grid of sampling points within the City of Palmerton.

4. Results

Figures 2-9 give 'post plots' showing metal concentrations (in mg kg^{-1}) measured at the various sample points for the top 15 cm. (in some cases, the measurements are averages from several analyses. For ease of reading, the measurements are shown at the design points rather than at the actual sample points). These plots make evident the high concentrations between and near the two smelter locations and the steady decline in concentrations as distance from the smelters increases.

Table I gives the measured metal concentrations for the ten duplicate pairs of samples, the absolute values D of pair differences of the original measurements in mg kg^{-1} and the absolute values L of pair differences of the logarithms of the measurements. Note that the sample variance computed on a pair of observations is $s^2 = D^2/2$. Hence, changes in D indicate how sample standard deviation is changing, from pair to pair. In Table I, it is obvious that, for each metal, D increases as the

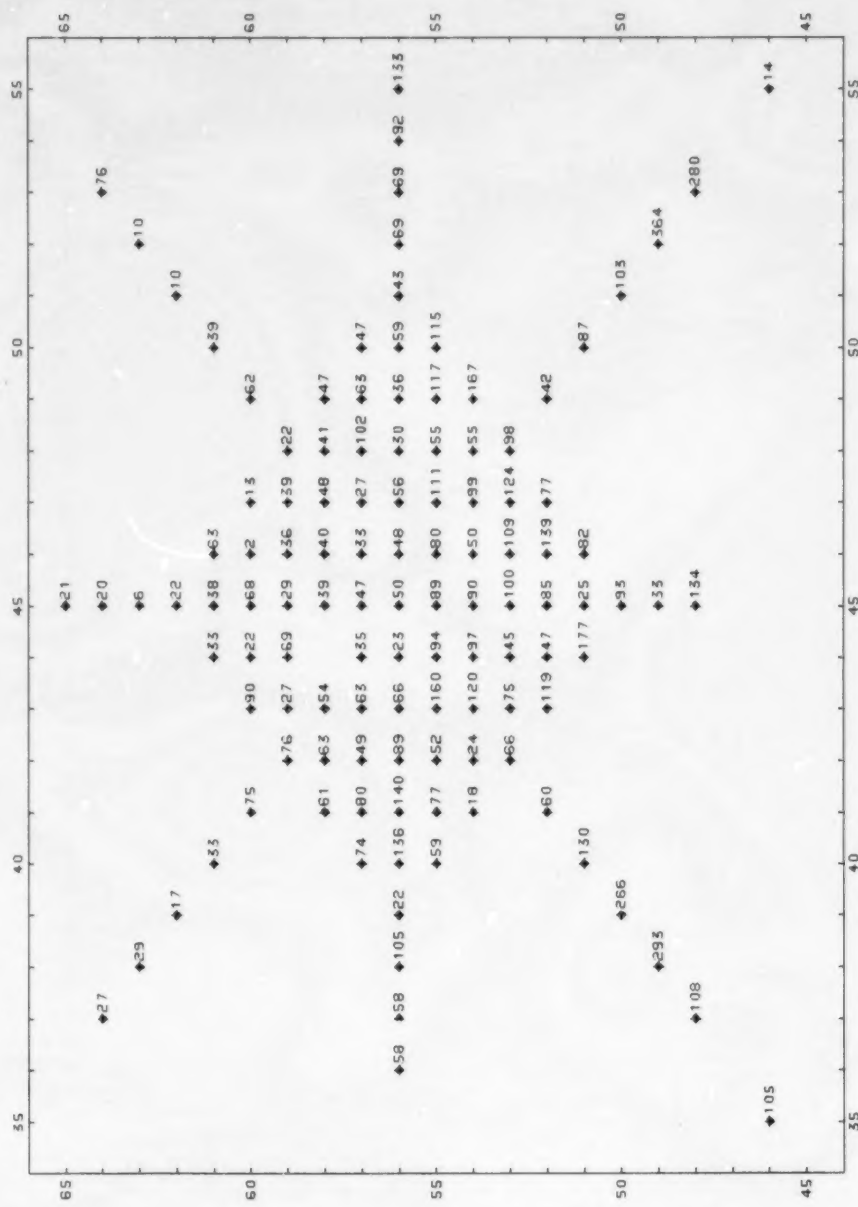


Fig. 2. Cadmium concentrations (mg kg^{-1}) near center of sample design.

PLOT NO. 1
DATE 08/26/86
TIME 12:56:29
PALMERTON SOIL MEASUREMENTS
CADMIUM CONCENTRATIONS

Fig. 3. Cadmium concentrations (mg kg⁻¹) along transects.

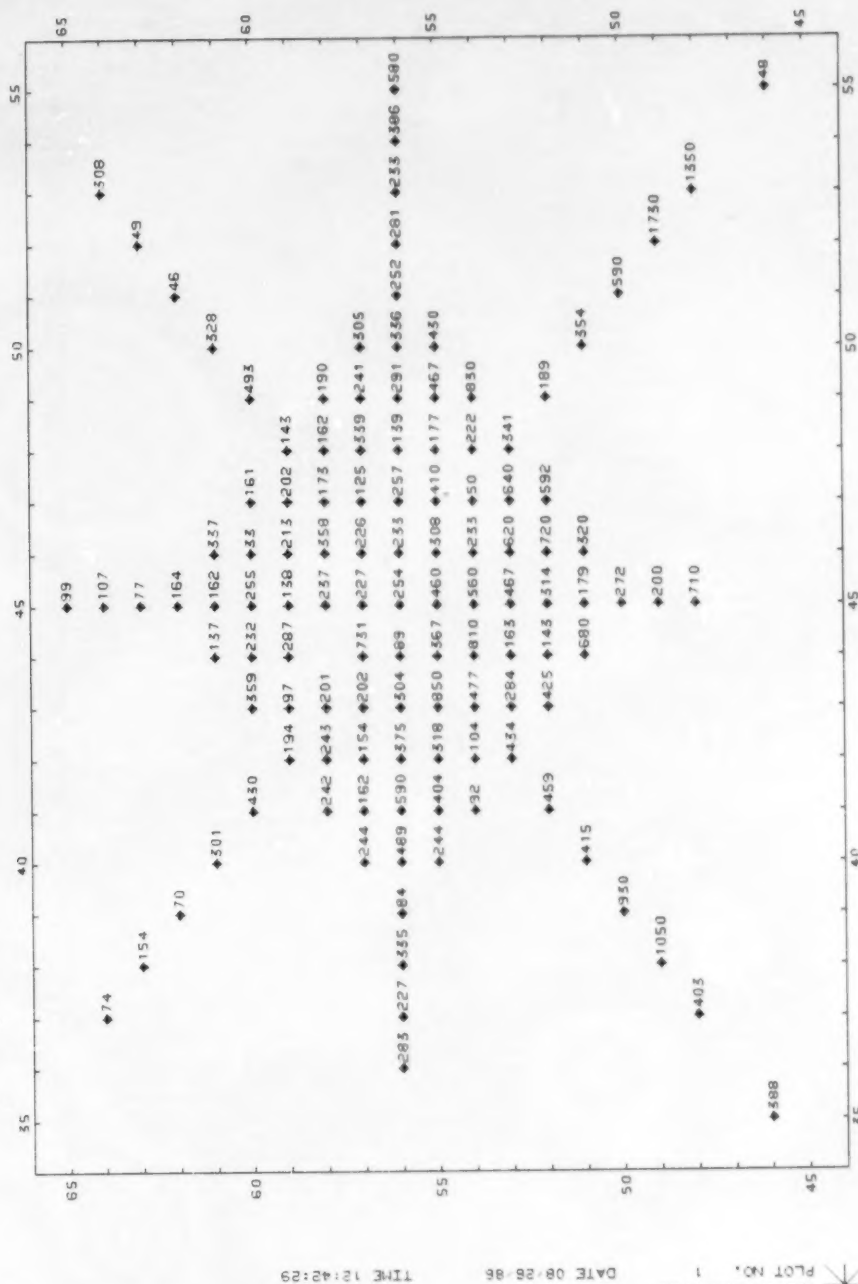
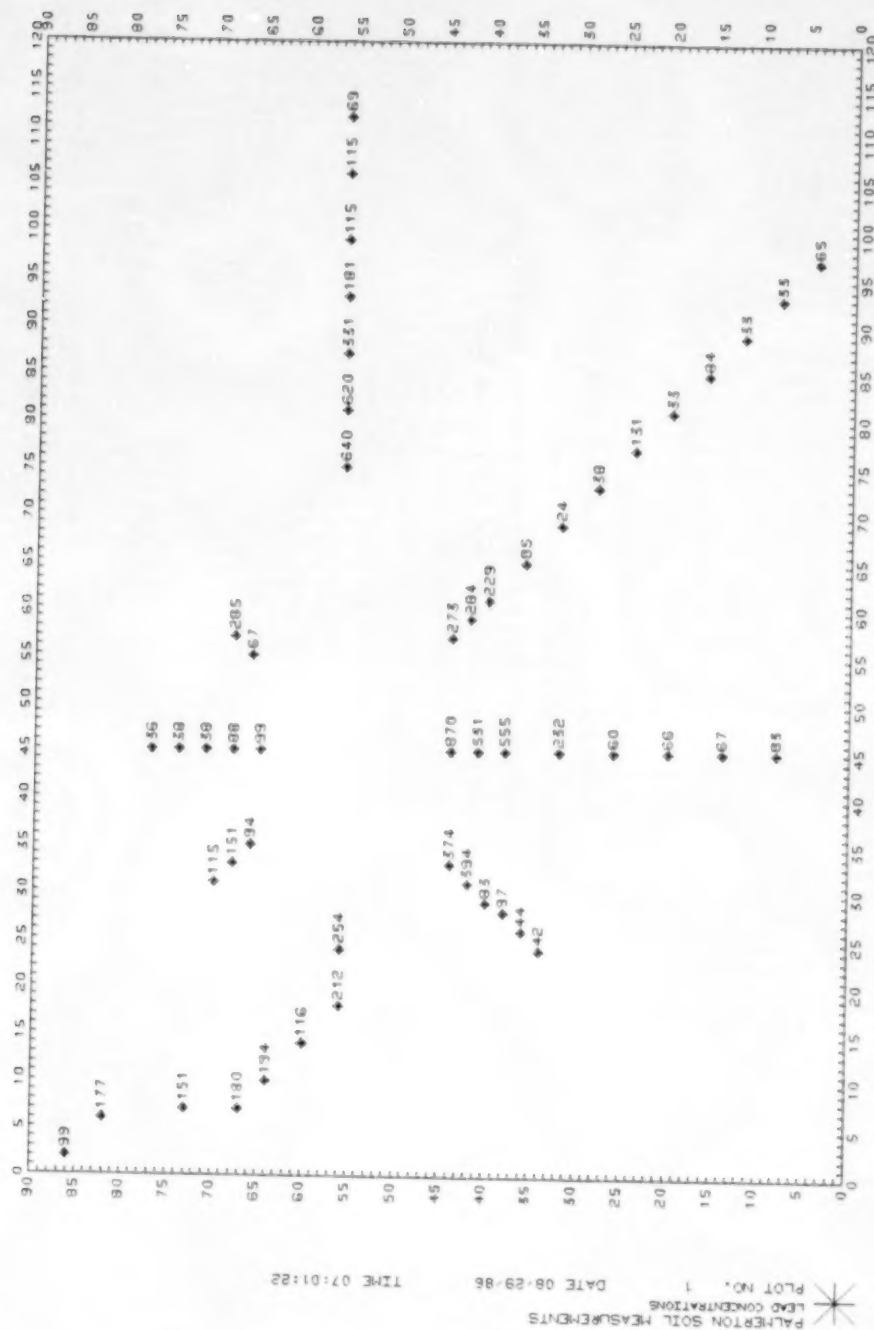
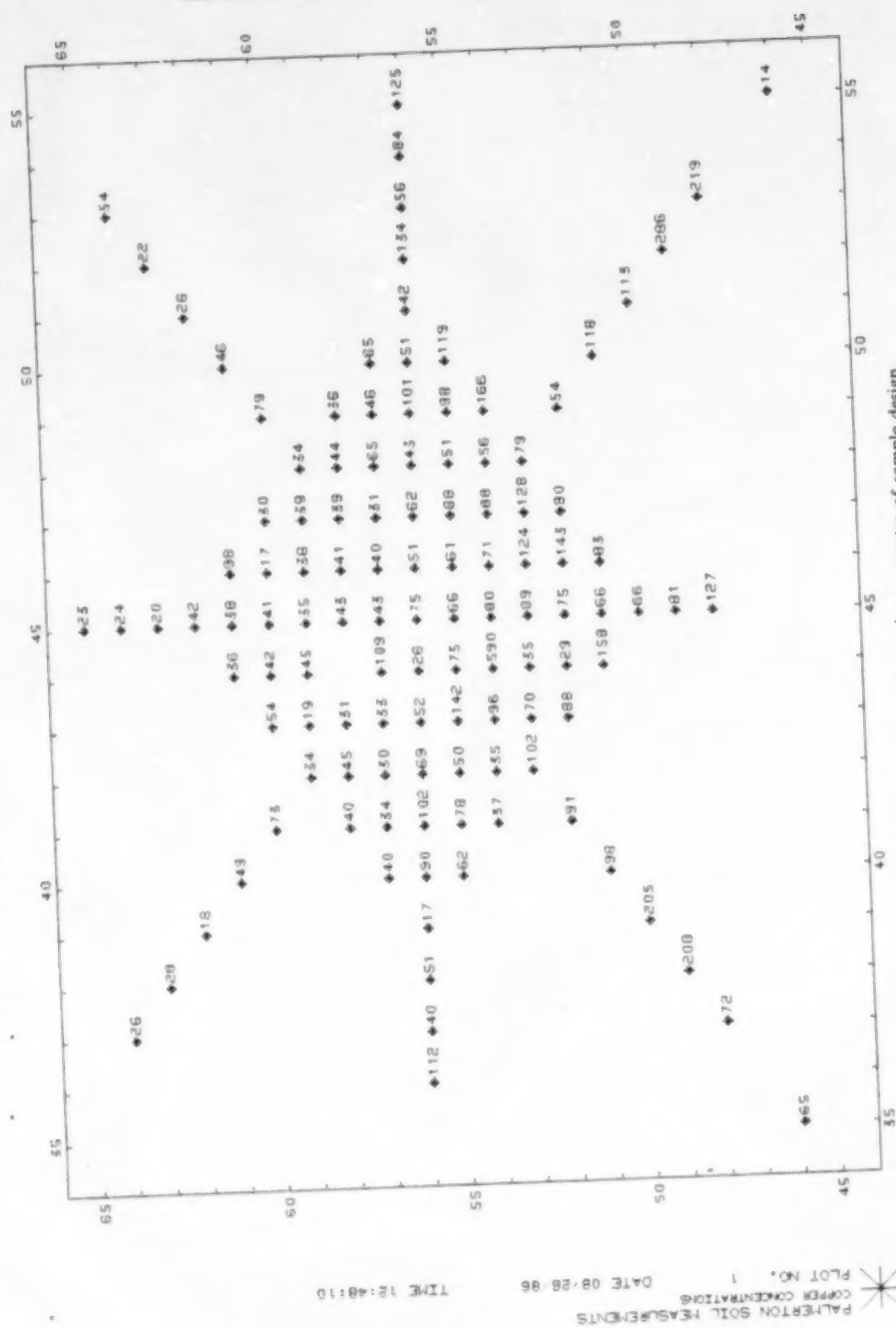
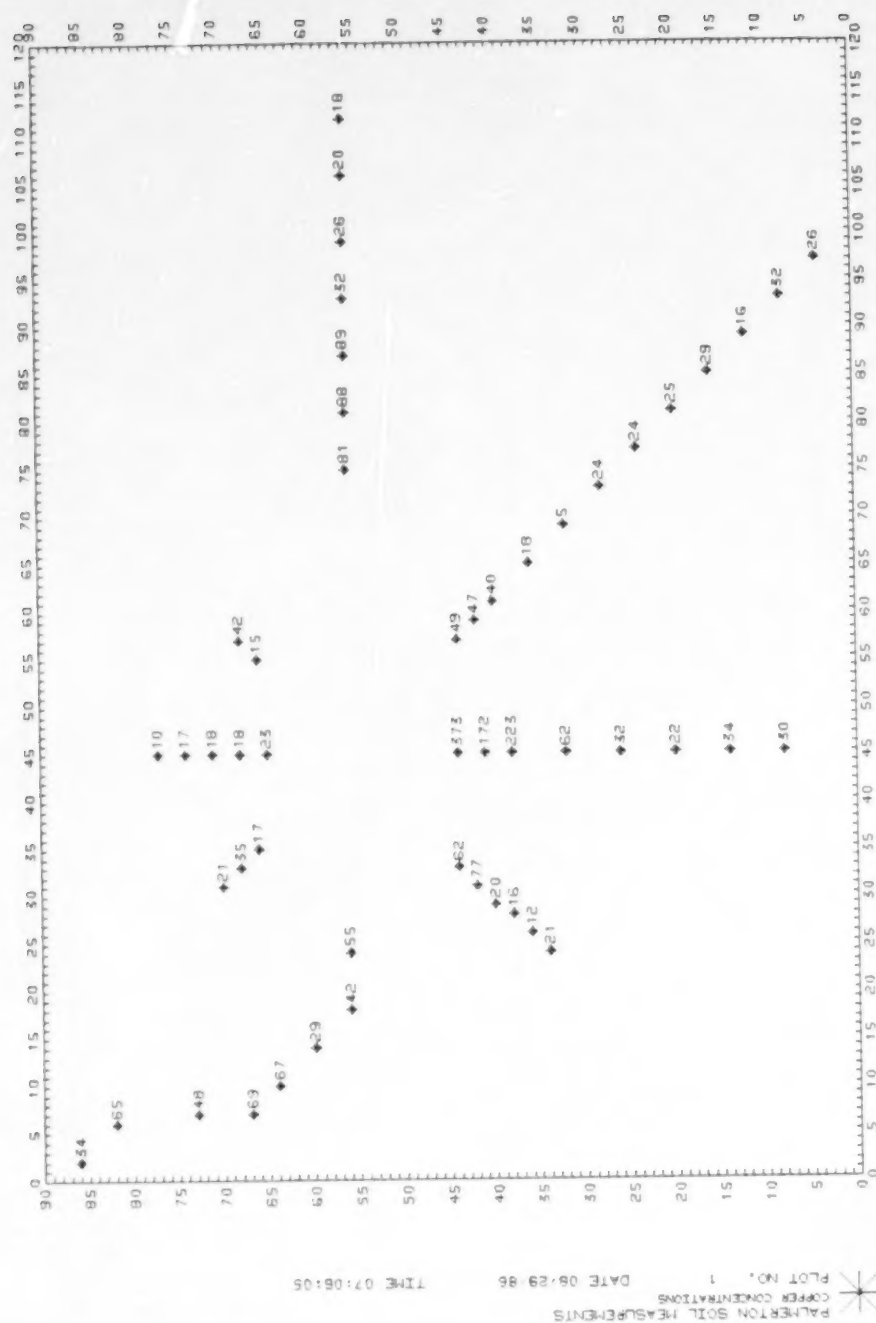


Fig. 4. Lead concentrations (mg kg⁻¹) near center of sample design.

Fig. 5. Lead concentrations (mg kg⁻¹) along transects.



Fig. 7. Copper concentrations (mg kg⁻¹) along transects.

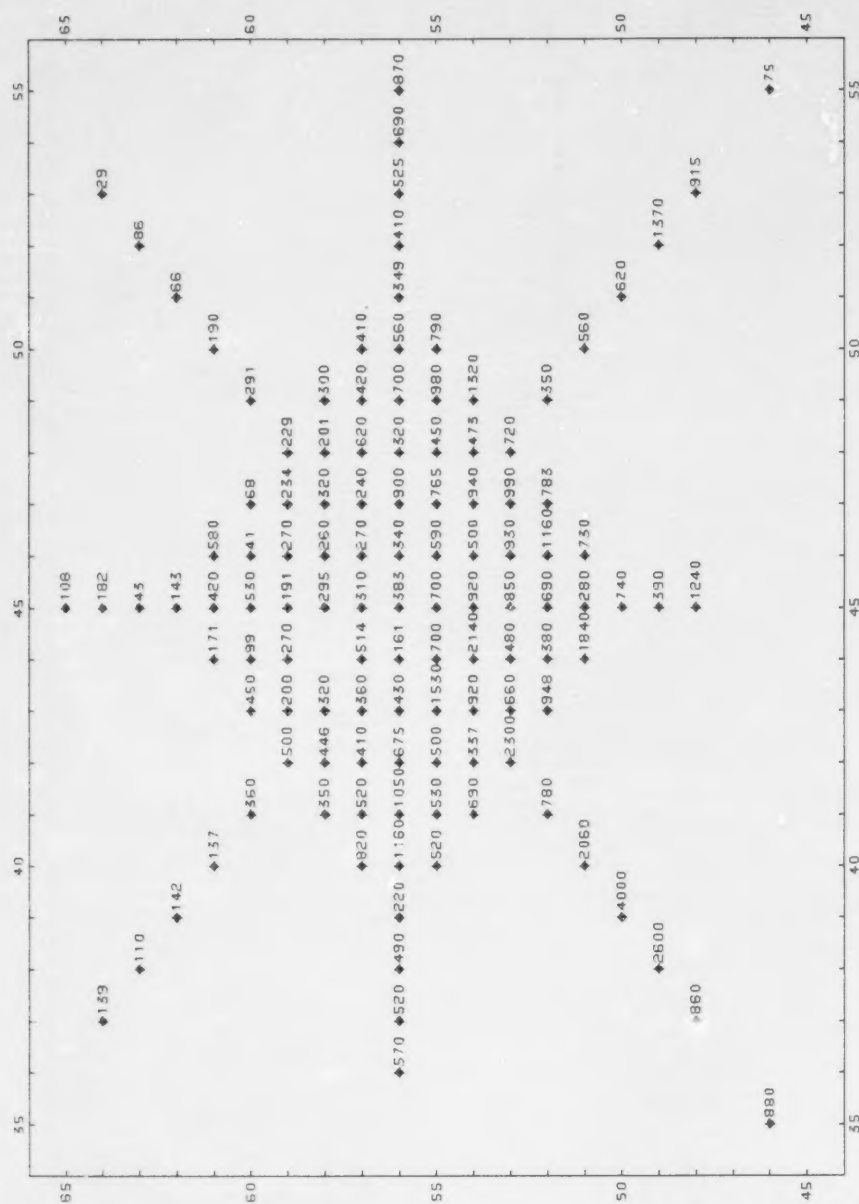


Fig. 8. Zinc concentrations (10 mg kg⁻¹) near center of sample design.

TIME 12:53:08

DATE 08/26/86

PALMERTON SOIL MEASUREMENTS
ZINC CONCENTRATIONS
PLOT NO. 1

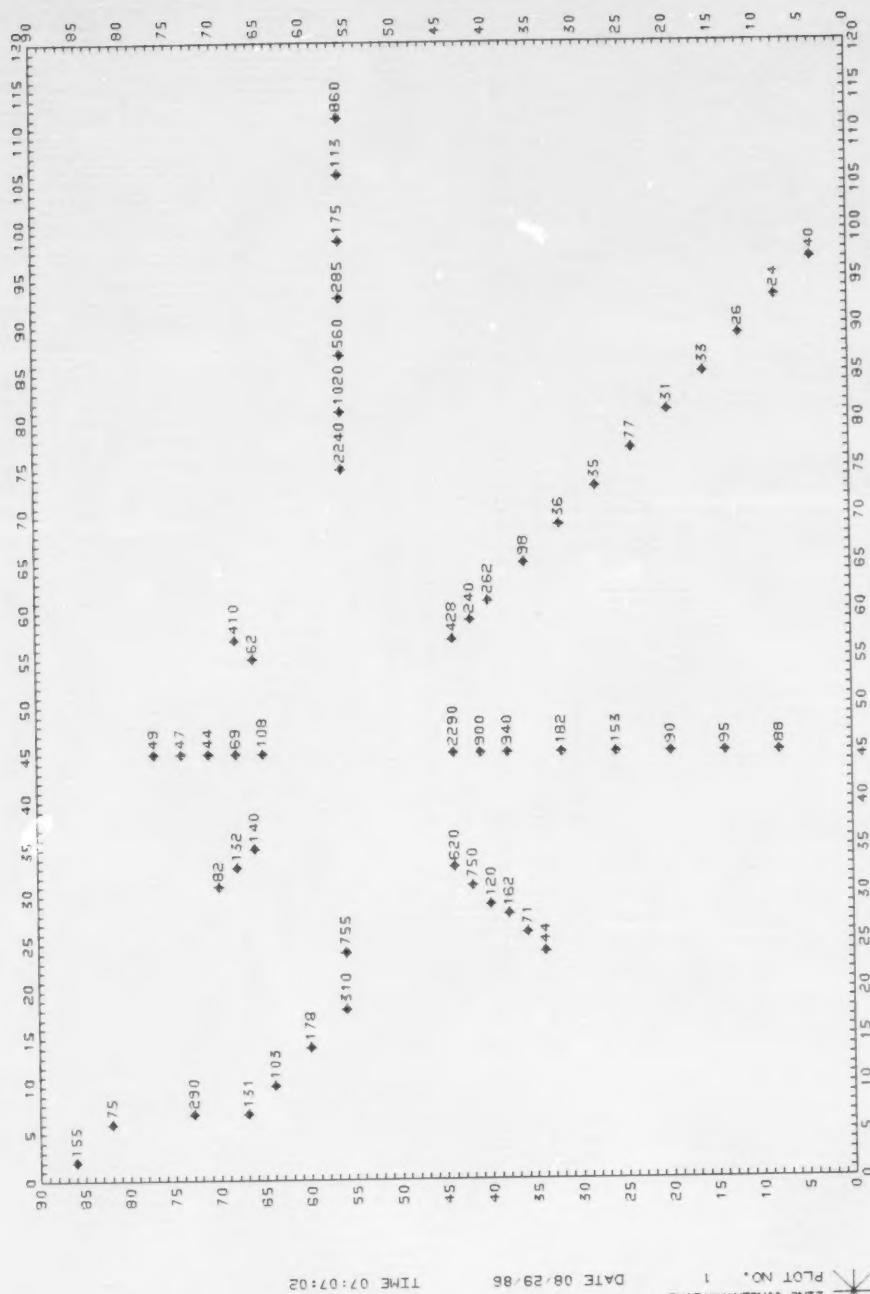
Fig. 9. Zinc concentrations (10 mg kg^{-1}) along transects.

TABLE I
Results from duplicate samples

Point	Cadmium		D^a		L^b	Lead		D	L
AK26	8.85	7.35	1.50		0.186	208	180	28	0.145
AO30	22.30	8.36	13.84		0.970	128	104	24	0.207
BP30	86.30	63.10	23.20		0.313	478	382	56	0.224
BQ33	58.20	40.30	17.90		0.368	221	87	134	0.932
BQ34	100.4	77.10	23.30		0.264	400	318	72	0.229
BT32	39.50	37.90	2.40		0.041	275	199	76	0.323
BT46	172.0	144.0	28.00		0.178	960	780	180	0.208
BY29	43.70	34.20	9.50		0.245	410	246	164	0.510
DL78	2.76	2.71	0.05		0.018	33.3	33.0	0.3	0.104
DP34	<u>68.10</u>	<u>61.70</u>	<u>6.40</u>		<u>0.098</u>	<u>201</u>	<u>161</u>	<u>40</u>	0.222
Median:			11.67		0.216			64	0.215
$s^2 = \Sigma L_i^2 / 20 = 0.0691$						$s^2 = \Sigma L_i^2 / 20 = 0.0751$			
Copper						Zinc			
AK26	68.0	65.0	3.0		0.045	1110	950	260	0.156
AO30	39.0	18.0	21.0		0.773	2700	860	1840	1.144
BP30	75.0	70.0	5.0		0.069	4200	3000	1200	0.336
BQ33	38.9	21.9	17.0		0.574	5100	3100	2000	0.498
BQ34	72.0	66.0	6.0		0.087	7800	5700	2100	0.314
BT32	46.0	40.4	5.6		0.130	3100	2800	300	0.102
BT46	393.0	352.0	41.0		0.110	27000.	18800.	8200	0.362
BY29	58.0	33.7	24.3		0.543	2030	1760	270	0.143
DL78	16.8	15.0	1.8		0.113	270	242	28	0.109
DP34	<u>33.2</u>	<u>30.9</u>	<u>2.3</u>		<u>0.072</u>	<u>3000</u>	<u>2700</u>	<u>300</u>	<u>0.105</u>
Median:			5.8		0.112			750	0.235
$s^2 = \Sigma L_i^2 / 20 = 0.0642$						$s^2 = \Sigma L_i^2 / 20 = 0.0989$			

^a D = absolute pair difference.

^b L = absolute pair difference of the log-transformed measurements.

average of the pair increases. However, the absolute difference L between the logarithms of the two observations in a pair, does not appear to be related to the average value of the pair. Therefore, the logarithmic transformation, $\ln$ (metal concentration), does appear to stabilize the standard deviation and variance. It is interesting to note that the estimated variance between log-transformed observations within duplicate pairs, $s^2 = \Sigma L_i^2 / 20$, is essentially the same for all four metals. This is what it should be once the effect of the size of the measurements is removed by the transformation since all four metal measurements are subject to the same kinds of short-range variation in deposition and errors in sample handling and analysis. (Investigation of the transformation, $\ln$ (metal concentration - m), where m was a positive number slightly smaller than the minimum observed concentration, was performed and it was found that the subtraction of m tended to over-correct the data).

TABLE II
Descriptive statistics for entire data set ($N=177$)

Statistics	Original observations (mg kg ⁻¹) Value	Log transforms Value
<i>Cadmium</i>		
Minimum	2.14	0.76
Maximum	364	5.90
Median	54.55	4.00
Mean	65.32	3.75
Std. Deviation	57.70	1.05
Coef. of Variation	88.33	28.03
Skewness	4.02	.55
Kurtosis	41.21	2.72
<i>Lead</i>		
Minimum	24.2	3.19
Maximum	1730	7.46
Median	242	5.49
Mean	299.3	5.38
Std. Deviation	247.0	0.85
Coef. of Variation	82.52	15.89
Skewness	4.45	0.16
Kurtosis	10.17	2.72
<i>Copper</i>		
Minimum	5.10	1.63
Maximum	590.0	6.38
Median	50.6	3.92
Mean	67.93	3.94
Std. Deviation	65.46	0.73
Coef. of Variation	96.37	18.54
Skewness	16.28	0.03
Kurtosis	27.37	3.43
<i>Zinc</i>		
Minimum	236.0	5.46
Maximum	40000	10.60
Median	4100	8.32
Mean	5450.6	8.12
Std. Deviation	5484.0	1.08
Coef. of Variation	100.6	13.3
Skewness	6.58	0.23
Kurtosis	12.98	2.63

For reference note that the skewness and kurtosis of a normal distribution are 0 and 3, respectively.

Descriptive statistics were calculated for the metal concentration measurements (for the top 15 cm) and for their log-transforms. (For sample points with splits, duplicates, or 2.5 cm subdivisions, the average of the measurements from the point is

TABLE III
Descriptive statistics for grid data ($N=100$)

Statistics	Original observations (mg kg^{-1}) Value	Log transforms Value
<i>Cadmium</i>		
Minimum	2.14	0.76
Maximum	177.0	5.18
Median	60.3	4.10
Mean	66.67	4.01
Std. Deviation	37.21	0.69
Coef. of Variation	55.81	17.2
Skewness	0.57	1.89
Kurtosis	3.10	7.09
<i>Lead</i>		
Minimum	32.5	3.48
Maximum	850.0	6.75
Median	256.0	5.55
Mean	311.1	5.57
Std. Deviation	180.7	0.61
Coef. of Variation	58.07	11.0
Skewness	1.20	0.24
Kurtosis	3.79	3.62
<i>Copper</i>		
Minimum	17.17	2.84
Maximum	590.0	6.38
Median	53.55	3.98
Mean	69.50	4.06
Std. Deviation	62.17	0.56
Coef. of Variation	89.46	13.87
Skewness	35.3	0.34
Kurtosis	49.3	4.57
<i>Zinc</i>		
Minimum	410.0	6.01
Maximum	23000	10.04
Median	4850	8.49
Mean	5833.3	8.43
Std. Deviation	4293.8	0.74
Coef. of Variation	73.61	8.81
Skewness	3.44	0.33
Kurtosis	7.13	4.12

employed). Table II gives the descriptive statistics for entire set of 177 points from which information was obtained. Table III gives the descriptive statistics for the set of 100 sample points consisting of 84 points on the square grid in the central diamond

TABLE IV
Descriptive statistics for reduced grid data ($N=69$)

Statistics	Original observations (mg kg ⁻¹) Value	Log transforms Value
<i>Cadmium</i>		
Minimum	6.18	1.82
Maximum	167.0	5.12
Median	58.7	4.07
Mean	66.44	4.03
Std. Deviation	36.47	0.62
Coef. of Variation	54.88	15.36
Skewness	0.54	0.51
Kurtosis	2.84	3.97
<i>Lead</i>		
Minimum	49.9	3.91
Maximum	850.0	6.75
Median	272.0	5.61
Mean	306.2	5.57
Std. Deviation	176.4	0.58
Coef. of Variation	57.62	10.33
Skewness	1.48	0.05
Kurtosis	4.18	3.01
<i>Copper</i>		
Minimum	18.8	2.93
Maximum	166.0	5.11
Median	52.0	3.95
Mean	63.75	4.03
Std. Deviation	32.57	0.49
Coef. of Variation	51.09	12.25
Skewness	1.02	0.01
Kurtosis	3.40	2.30
<i>Zinc</i>		
Minimum	430	6.06
Maximum	20600	9.93
Median	4300	8.37
Mean	5341.7	8.36
Std. Deviation	3637.0	0.70
Coef. of Variation	68.09	8.42
Skewness	2.45	0.26
Kurtosis	6.34	3.70

(Figure 1) and 2 points closest to the central diamond from each of the 8 transects. Table IV gives the statistics for the 69 points remaining from the 100 used in Table III after 31 points were removed because their site description forms indicated that they represented samples on recent fill, severely eroded soil, or recent cuts, and thereby might not give an accurate indication of the metal concentrations in the soil in their immediate vicinities. The decision to keep or exclude a point in this set was not based on measurements of the metal concentrations obtained at the point.

TABLE V
Correlations between log-transformed metal concentrations

	ln (Cd)	ln (Pb)	ln (Cu)	ln (Zn)
(N=177)				
ln (Cd)	1.000	0.889	0.766	0.901
ln (Pb)		1.000	0.883	0.840
ln (Cu)			1.000	0.797
ln (Zn)				1.000
(N=69)				
ln (Cd)	1.000	0.706	0.789	0.921
ln (Pb)		1.000	0.791	0.656
ln (Cu)			1.000	0.795
ln (Zn)				1.000

The product moment sample correlations between the log-transformed metal measurements are given in Table V for the entire data set ($N=177$) and for the reduced set of grid points ($N=69$). As would be expected, these correlations are all highly positive. The correlations between the log transforms of Cd, Cu, and Zn are virtually the same in the reduced data set as in the full data set, while the correlations with the log-transformed lead concentrations change. The reason for this can be seen by looking at the plots of metal concentrations, Figures 2-9, which show a decreasing trend in cadmium, copper, and zinc concentrations from south to north on the Palmerton grid, but a less clearly defined trend for lead concentrations. However, over the transects one observes the same basic trends in concentrations for all metals.

TABLE VI
Results from splits

Site	Cadmium ^a		D	L	Lead ^a		D	L
DD70	4.17	4.13	0.04	0.010	33.4	32.1	1.3	0.040
BV35	116	106	10.0	0.090	430	389	41.0	0.100
CB34	73.8	63.1	10.7	0.157	244	221	23.0	0.099
BT70	7.34	7.32	0.02	0.003	72	60	12.0	0.182
AY34	88.0	81.9	6.1	0.072	275	233	42	0.166
BT35	95.1	83.1	12.0	0.135	474	446	28.0	0.061
DP82	2.50	2.43	0.07	0.028	33.5	32.7	0.8	0.024
CB42	280	279	1.0	0.004	1360	1340	20.0	0.015
BR34	68.7	63.3	5.4	0.082	313	294	19.0	0.063
CD24	9.1	8.99	0.11	0.012	67	67	0	0.000
$s^2 = \Sigma L_i^2 / 20 = 0.0032$					$s^2 = \Sigma L_i^2 / 20 = 0.0045$			

^a Measurements in mg kg⁻¹

TABLE VII
Results from individual cores

Point	CORES					V*
	S	W	N	E	Range	
Cadmium (mg kg ⁻¹)						
BO29	30.7	<u>5.0</u>	<u>66.3</u>	29.0	61.3	1.195
BR33	56.7	66.8	<u>25.8</u>	<u>104.3</u>	78.5	0.339
BR38	104.0	<u>147.0</u>	124.0	<u>100.6</u>	46.4	0.031
BS33	42.9	<u>21.2</u>	30.5	<u>45.5</u>	24.3	0.124
BU33	29.4	<u>24.1</u>	31.0	<u>48.6</u>	24.5	0.877
BU38	64.0	<u>33.1</u>	82.3	<u>127.0</u>	93.9	0.316
BX30	56.3	<u>35.9</u>	46.2	<u>111.0</u>	75.1	0.234
CF46	51.5	<u>97.1</u>	<u>46.7</u>	79.0	50.4	0.121
CI34	<u>112.0</u>	167.0	158.0	<u>196.0</u>	84.0	0.055
CJ50	0.75	47.4	115.0	52.0	114.25	5.159

* Sample variances on ln(Cd) over 4 cores at each site.

Extreme core values at each sample point are underlined.

Pooled sample variance based on first nine sample points: 0.3659.

Pooled sample variance based on all ten sample points: 0.8453.

Lead (mg kg ⁻¹)						
BO29	218	220	238	<u>528</u>	310	0.183
BR33	134	197	<u>99</u>	<u>377</u>	278	0.333
BR38	274	500	<u>513</u>	413	239	0.084
BS33	<u>1980</u>	<u>105</u>	306	531	1875	1.493
BU33	201	235	<u>284</u>	<u>183</u>	101	0.037
BU38	398	<u>308</u>	730	<u>930</u>	622	0.265
BX30	397	334	<u>241</u>	<u>1000</u>	759	0.372
CF46	<u>132</u>	<u>385</u>	197	<u>377</u>	253	0.273
CI34	<u>1000</u>	880	<u>750</u>	830	250	0.014
CJ50	<u>10.90</u>	201	<u>527</u>	176	516.1	2.785

* Sample variance of ln(Pb) over four cores.

Pooled sample variance over first nine samples: 0.339.

Pooled sample variance over all ten samples: 0.584.

The possible cause of this anomaly is discussed below in a paragraph on spatial structure. The results from the splits (i.e., subsamples taken from a sample after sieving and mixing) for cadmium and lead are given in Table VI. The results for copper and zinc are similar. These results indicate that very little of the variation observed between duplicates in Table I was due to subsampling and analytical errors. The variance of log-transformed measurements from the pairs of splits is about one-twentieth that observed between log-transformed measurements in the duplicate pairs.

The measurements of cadmium and lead in individual cores coming from ten of the sample points are given in Table VII. Large variation between cores is evident. This large variation between cores is a major source of the variation between

duplicate pairs observed in Table I. Hence, to reduce the variation between duplicate pairs and the short range variability of measurements one must take, and composite, more cores at each sample point (see Starks, 1986). (The usual formula for variance of a mean showing the effect of increasing sample size does not apply here since the logarithm of an average (composite) concentration does not equal the average of the logarithms of individual core concentrations).

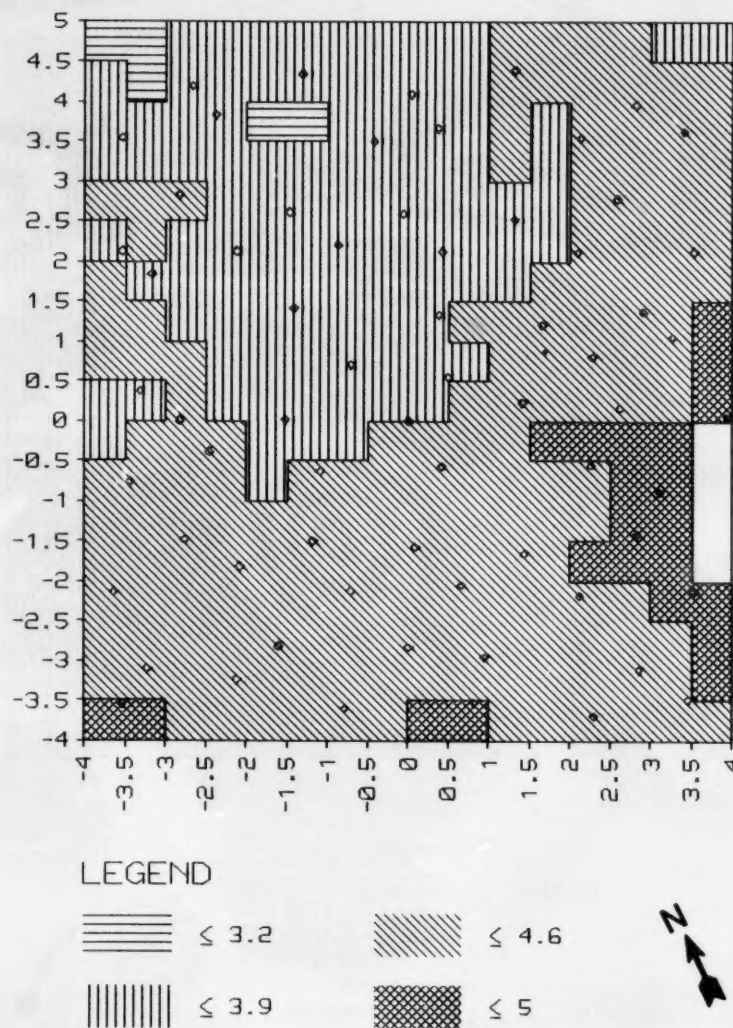


Fig. 10. Block kriging results for log-transformed concentrations of cadmium in grid area (block size: 200' x 200'). Actual sample point locations are indicated by the symbol $\odot$. Note: $\ln(24.52) = 3.2$, $\ln(49.40) = 3.9$, $\ln(99.48) = 4.6$, and $\ln(148.40) = 5.0$. Map units are 122 m (400').

The analysis of spatial structure involves the estimation of a generalized covariance function and an order k for the intrinsic random function representing each log-transformed metal concentration in the soil (see Delfiner, 1976, for a discussion of generalized covariance functions; and Gilbert and Simpson, 1985, and Flatman and Yfantis, 1984, for discussions of the use of geostatistics in soil sampling). The estimation was based on information from the 69 data points within the central

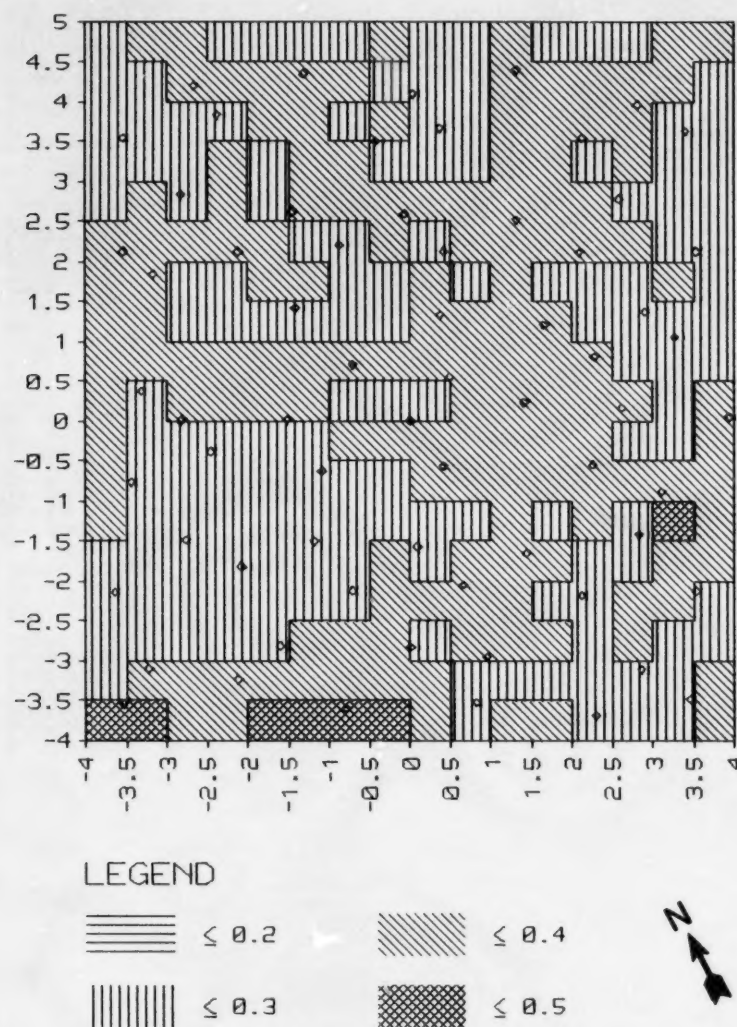


Fig. 11. Standard errors for block kriging estimates of log-transformed cadmium concentrations. Actual sample point locations are indicated by the symbol $\diamond$.

diamond that were not excluded because of their site description. White-noise and spherical models for the generalized covariance function were tried. Application of the estimation procedure based on cross-validation and response surface methods to the data produced white-noise models and $k=2$, for each of the four metals. The sills of the white-noise generalized covariance functions were 0.22 for Cd, 0.34 for Pb, 0.18 for Cu, and 0.21 for Zn. One might have expected the sills to have been the same (except for some experimental error variation), as those for Cd, Cu and Zn seem to be, since they have all been deposited by the same process. However, the sill for lead is much larger than for the other three metals, and, in addition, there is less evidence of drift over the grid points. The white-noise model with $k=0$, with estimated sill 0.30, gives very nearly as good a fit of the $\ln(\text{Pb})$ data as it does with $k=2$. In addition, the estimated sill (when $k=2$), 0.34, is the same value as the sample variance of the 69 $\ln(\text{Pb})$ measurements which again indicates little evidence of drift. A possible reason for the difference in spatial structure between lead and the other metals is that most of the lead emissions over the latter years of operation came from the West Plant while most of the other metal emissions came from the East Plant. During the last 10 yr of operation, 1970-1979, before the 1980 plant modifications, the percentages of metal emissions coming from the East Plant were estimated to be 71% for Zn, 67% for Cd, 100% for Cu, and 37% for Pb with the remainder of the emissions coming from the West Plant (Silvestris, 1984).

The results of block kriging (block size: 61 m square (200' $\times$ 200')), over the central diamond area of Figure 1, to estimate cadmium concentrations using the white-noise model and drift of degree 2 is given in Figure 10. A corresponding standard error plot, for these block kriging estimates is given in Figure 11.

5. Discussion of Procedures and Results

One of the first steps in the data analysis was to determine whether a variance stabilizing transformation was needed. It was obvious from the nature of the metal deposition and from the post plots of the data that drift is present (i.e., the expected value for metal concentration in the soils is not the same at all locations in the study area). If the variance of a random function changes when the mean changes, then sills, nugget effects, and spatial correlations change from location to location and no one generalized covariance function will apply to the data. However if the random function can be transformed so as to stabilize its second moment properties, then one generalized covariance function will apply and can be estimated from the sample data. One cannot, as is commonly advocated, determine the appropriate transformation by looking at the nature of the empirical distribution of the sample data because the empirical distribution of the sample data is just an artifact of the choice of sampling locations and the nature of the drift over the sampled region. It is the distribution of the random variable defined by the random function at each point in the sample space that is of interest. The empirical distribution of the sample data will give information concerning the distribution of these random variables only if

the sample points are uniformly distributed over the region and there is no drift. By looking at differences and sums of measurements at nearby pairs of sampling locations, one observes how second moment properties (variance and covariance) are changing as first moment properties change due to the drift. For the Palmerton data set, there were only ten duplicate pairs, which is certainly a small number for determining a variance stabilizing transformation, however the excellent agreement between variances of the transformed data over the four metals added additional evidence in favor of the log-transformation that was chosen.

While the estimated spatial structure model for each metal is one with a white-noise generalized covariance function, this really means that the range of influence is not larger than the spacing between closest sample points (namely, 122 m). The sills for three of the white-noise models are about 0.20, and the one for lead is 0.34. However, the variances obtained by looking at the differences of the ten duplicate sample results are near 0.08. This indicates that there is a range of influence larger than 0.5 m (the distance between duplicate pairs) and smaller than 122 m. In future sampling, it will be useful to take some samples at shorter spacings (say 12 m apart) so as to get an estimate of this range of influence and thereby improve the kriging variances in block kriging of blocks near sample points.

6. Notice

Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through Cooperative Agreement CR 812189-01 to the Environmental Research Center of the University of Nevada, Las Vegas, it has not been subject to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Acknowledgment

The authors thank Mr. Ed Shoener, U.S. Environmental Protection Agency, Region III, for his suggestion as to the probable reason for the differences in spatial structure and correlation patterns detected between lead and the other three metals.

References

- Delfiner, P.: 1976, 'Linear Estimation of Non-stationary Spatial Phenomena', in M. Guarascio, M. David, and C. Huijbregts (eds.), *Advanced Geostatistics in the Mining Industry*, D. Reidel Publ. Co., Dordrecht, pp. 49-68.
- Flatman, G. T. and Yfantis, E. A.: 1984, 'Geostatistical Strategy for Soil Sampling: The Survey and the Census', *Environmental Monitoring and Assessment* 4, 335-349.
- Gilbert, R. O. and Simpson, J. C.: 1985, 'Kriging for Estimating Spatial Pattern of Contaminants: Potential and Problems', *Environmental Monitoring and Assessment* 5, 113-136.
- Silvestris, M.: 1984, Untitled letter to E. Shoener dated September 13, 1984.
- Starks, T. H.: 1986, 'Determination of Support in Soil Sampling', *Mathematical Geology* 18, 529-537.

Starks, T. H. and Sparks, A. R.: 1987, 'On the Estimation of the Generalized Covariance Function: II. A Response Surface Approach', *Mathematical Geology* **19**, 765-779.

Starks, T. H., Brown, K. W., and Fisher, N.: 1986, 'Preliminary Monitoring Design for Metal Pollution in Palmerton, Pennsylvania', in C. L. Perket (ed.), *Quality Control in Remedial Site Investigation: Hazardous and Industrial Waste Testing* **5**, ASTM STP 925, American Society for Testing and Materials, Philadelphia, pp. 57-66.